

DOCUMENT RESUME

ED 178 609

TM 009 851

AUTHOR Archambault, Francis X, Jr.; And Others
 TITLE The Classroom Activities Log: A Personalized Record of a Child's School Day as Recorded by His/Her Teacher.
 INSTITUTION Abt Associates, Inc. Cambridge, Mass.
 SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
 PUB DATE Apr 79
 CONTRACT NIE-400-75-0074
 NOTE 29p.; Paper presented at the annual meeting of the National Council on Measurement in Education (San Francisco, CA, April 9-11, 1979)
 EDRS PRICE MF01/PC02 Plus Postage.
 DESCRIPTORS Bilingual Education; Class Size; Compensatory Education Programs; *Data Collection; Elementary Secondary Education; Group Instruction; Intellectual Disciplines; *Measurement Techniques; Observation; Records (Forms); *Student Records; Teacher Aides; *Teacher Attitudes; Test Construction; Test Reliability; Test Validity; *Time Factors (Learning)
 IDENTIFIERS *Classroom Activities Log; *Instructional Time

ABSTRACT

The development and validation of an instrument designed to measure the quantity of types of instruction delivered to individual students is described. The Classroom Activities Log (CAL) is a teacher-completed record of students' instructional experiences, and may be used to study educational equity questions posed by Elementary Secondary Education Act Title I mandates. Some of the features of CAL are: (1) student focus; (2) limited number of variables; (3) minimal teacher training and costs; (4) availability of data on individual students' daily activities; and (5) ease of scaling. The school day is divided into ten-minute periods, and thirteen content areas are quantified: language arts; English as a second language; mathematics; social studies; science; fine and applied arts, home economics, shop, or foreign languages; physical education, health and safety; career education; cultural enrichment; special services; recess, lunch or unoccupied; classroom management; and other. Information is also provided on size of the instructional group, type of adult involved in the instruction; bilingual instruction; and compensatory education programs. It is concluded that reliability estimates (based on correlations between teacher and observer ratings) are acceptable, and that there is a certain degree of construct validity. (MH)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED178609

TM009 851

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE CLASSROOM ACTIVITIES LOG:
A PERSONALIZED RECORD OF A CHILD'S SCHOOL DAY
AS RECORDED BY HIS/HER TEACHER

Francis X. Archambault, Jr.
University of Connecticut

Nancy L. Ames
Glenn Takata
Abt Associates, Inc.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Francis X Archambault, Jr.

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

National Council for Measurement in Education, Annual Meeting

San Francisco, California
April 1979

Author's Note: This paper is based in part on research performed by Abt Associates, Inc., under contract No. NIE-100-75-0074 to the National Institute of Education. The statements in this paper do not reflect any official position of the Institute.

This paper describes the development and validation of an instrument designed to measure the quantity of various types of instruction delivered to individual students. Interest in this area of research may be traced to Carroll's (1963) model of school learning in which "time spent in learning" played a key role. Two basic research thrusts have evolved from Carroll's work. One of these has focused on the effects of various amounts of schooling (or instruction) on student achievement. Wiley and Harnischfeger (1974), for example, have done considerable work in this area and have published results indicating that amount of schooling does have an effect on student performance. Moreover, they claim that their results explode the "no school effects" myth of Coleman, et al. (1966), Jencks, et al. (1972), Mosteller and Moynihan (1972), and others. Bloom (1976), Berliner (1978), Karweit (1976) and Fisher, Filby and Marliave (1978), to mention only a few, are also actively involved in research on the effects of quantity of instruction.

A second set of studies has focused on the quantity of instruction as an outcome measure unto itself. Studies addressing the educational equity of various Federal programs have taken this approach. The National Institute of Education sponsored studies of ESEA Title I, for example, have compared the amounts of various types of instruction received by Title I and non-Title I students to determine whether Federal funds were being used to provide services not available through use of local or state funds. (See Vanecko et al., 1977 or Frechtling and Nyitray, 1978, for example)

Despite the emphasis on and need for quantity of instruction data, a very interesting and knotty problem remains to be solved. This problem is best expressed in the form of a question: How can data on the quantity or amount of instruction delivered to or received by individual students be collected on a large enough scale, at reasonable cost, and with sufficient technical quality to support large scale studies of school effects or educational equity? This paper describes an instrument which appears to satisfy these conditions. This instrument, which is called the Classroom Activities Log (CAL), is a teacher-completed record of students' instructional experiences, and is based in large part on the Harinschfeger and Wiley (1977) school effects model.

Review of Research

In general, data on the quantity of instruction can be obtained in five ways: (1) directly from students; (2) through the use of trained observers; (3) through teacher reports; (4) through some combination of approaches (1), (2) and (3); or (5) from school records. By far the most popular of the approaches for quantifying classroom events is the classroom observation technique. Included under this general heading are all instruments which use trained observers to acquire data on teacher/student behavior. Well over 120 instruments of this kind were available when Rosenshine and Furst (1973) reviewed the literature, and more have been developed since. As one might expect, these observation techniques differ in a number of ways, including: (1) differences in recording procedures; (2) differences

in items, particularly with respect to the degree of inference required of the observer; and (3) differences in format. In addition to these, instruments also vary widely in quality, from carefully developed and tested instruments to carelessly-constructed, non-standardized techniques.

Despite their numbers, several aspects of such classroom observation instruments mitigate against their use in quantity of instruction studies. These may be summarized as follows:

1. Most classroom observation instruments (with notable exceptions) focus on teacher and not student behavior.
2. There appears to be a lack of agreement on the constructs to be observed and measured which often results in a shotgun approach to data collection intended "to cover all the bases."
3. There is an obvious need for independent observers which necessitates training to ensure that behaviors are perceived similarly and that ratings are consistent.
4. As a result, the costs of data collection often become excessive, particularly in large scale research.
5. To compensate for this, behavior samples are often obtained rather than observations over longer periods of time, such as whole school day.
6. Finally, many such instruments require elaborate rating or scoring mechanisms in order to construct scales.

Wiley and Harnischfeger (1974) avoided these problems in their early work by obtaining quantity of schooling data from school records. Using this approach, quantity of schooling was defined as the triple product of average daily attendance for a particular school, the number of hours per day, and the number of days per year the school met. As can be seen from this definition, quantity of schooling was a school level variable. Consequently, the variation in instruction provided to individual students was lost in any analyses of the data.

The Classroom Activities Log (CAL) supports student level analyses and overcomes the problems discussed above. Some important features of the CAL are the following:

- (1) it focuses on the student rather than the teacher;
- (2) it contains a limited number of variables which are justified by the recent quantity of schooling literature;
- (3) it can be completed by classroom teachers with a minimum of training and costs;
- (4) it provides complete data on individual students' experiences for a whole school day; and
- (5) it is easy to scale.

Further, the CAL is the only instrument with such flexibility successfully used in a large scale data collection effort.

The CAL provides information on what some researchers have referred to as "allocated time." That is, it provides a teacher estimate of the amount of time an individual student can devote to a particular content or instructional area. This time can be allocated by the student or by the teacher, although usually the latter, and sets an upper limit on the amount of learning that can take place. Some current research also addresses what is called "engaged time." According to Fisher (1978), "engagement refers to the involvement of the student in the learning actively under consideration" (p. 3). He goes on to say that student learning can occur only if there is some degree of engagement, and as a result, he argues that the distinction between engaged time and allocated time is critical. This is supported to some degree by small scale studies conducted by Fisher and others involved in the Beginning Teacher Evaluation Study (i.e., Berliner, 1978; Marliave, 1978; Fisher, Filby and Marliave, 1978).

Despite the plausibility of the arguments raised above, it should be noted that "engaged time" is collected by classroom observers. Further, these observers are required to make inferences of various kinds (e.g., is the student in question "paying attention" is determined through behavioral cues such as posture, eye contact with the material, and so forth). Possibly due to differences in reliability of the various measures, research described by Marliave (1978) reveals that allocated time may be as good a predictor of student achievement as engaged time. For these reasons, the collection of data on allocated time, as described below, and its use in studies of school effects and educational equity appears warranted.

The Development of the CAL

The CAL was developed to address a series of educational equity questions within the Congressionally mandated Title I Demonstration Study (Vanecko et al., 1978). The CAL divides the school day into ten minute intervals and is completed by teachers at convenient times throughout the day. Directions for the completion of the machine-scorable form are carefully detailed in a comprehensive teacher's manual (Abt Associates, 1977).

A portion of the CAL form is reproduced in Figure 1. As can be seen in the figure, the CAL provides quantity of instruction data for 13 specific content areas (see the "content area" dimension of the figure). It also provides information on the size of the group in which instruction was delivered ("Group Size") and the

A Portion of the CAL

+

type of adult directly involved in the delivery ("Adult Involved").¹ These latter two variables, described by Wiley and Harnischfeger in their school effects model, are refinements of the quantity of schooling notion. We refer to these two variables as measures of the "intensity" of instruction.

In addition to information on quantity and intensity of instruction, the CAL also describes whether or not the instruction was compensatory and/or bilingual. These two dimensions were necessary to make comparisons in the Title I Demonstration Study, but the CAL can be easily modified and used without them at no risk to its integrity.

The development of the CAL included a number of procedures. After reviewing the available research literature, a pilot version of the instrument was developed.² This early edition of the CAL was piloted in a number of small-scale field tests using teachers from different schools and communities. Perhaps the most important conclusions drawn from these trials was that teachers could record the quantity and intensity data for a single student, but that trying to do more than this would affect their teaching behavior. We therefore decided to prescribe that only one sampled student be observed on any given day.³

¹Due to space limitations definitions of the 13 "content area" categories as well as the categories for other dimensions of the CAL are not presented. Interested readers may find such definitions in the teachers manual.

²The authors would like to thank James Vanecko, David Wiley, Thomas Johnson, Dale Mann, Ann Milne, Iris Rotberg and Paul Hill for their useful comments during the development of the CAL.

³To insure representativeness in the Demonstration Study, we sampled both the student and the day.

The Reliability of the CAL

Of particular concern in the determination of the CAL reliability was the extent to which teachers, who should know best what a student does or experiences in the classroom, would be able to accurately capture these experiences on the CAL form. Field trials had indicated that because of their various regular activities teachers would most likely complete the form between classes or during breaks. As a result, the responses would be based on recall. The major issue, therefore, was the extent to which job pressures as well as lack of interest or commitment to data acquisition could affect the accuracy and completeness of this recall.

To check on this, a small sample of children⁴ was selected for independent observation. Observers followed the child throughout the day and did not have to depend on recall. On the other hand, they knew less about the child's daily activities and could not always determine the appropriate description in ambiguous situations. A comparison of the two sets of responses provides a way to determine the amount and kinds of error affecting the observations.

The estimate of CAL reliability is based on measures of the amount of correspondence between the observations from the teacher and from the outside observer. In addition, a more specific examination of typical kinds of mismatches allows for a determination of the amount and kinds of errors that arise. As was noted above, however,

⁴A total of 124 children, or approximately eight from each of the then fifteen demonstration districts, were used in the study. More than 76 observers were also used, each of whom was trained for this purpose.

the inaccuracies could be due to the observer as well as the teacher. The determination of the sources of error cannot be done analytically insofar as any error pattern could be generated by several different kinds of observer or teacher behavior. Discussion of these sources of error are, therefore, based mainly on knowledge about how people, especially teachers and observers, react to this type of instrument.

Among the many potential hypotheses about observational error, there is one type that will not be discussed. This is the class of biases that relate to the nature of the validation study itself. In any study of this kind, where two people are "checking" on each other, it is possible that the teachers and observers make extra efforts to give desirable results. Thus, teachers might be more attentive to the task than under normal conditions. As a result, there is the possibility that the data reported below represent best rather than normal efforts, making the estimates of reliability biased toward the high side. Since this issue will not be discussed further, the possibility of upward bias must be considered when examining the analytic findings.

Analytic Techniques

Two measures of correspondence were used for assessing the reliability of the five CAL dimensions. Both are based on the total number of ten minute blocks containing valid information. This measure of the length of the school day is determined by counting the number of content area blocks in which either observer provides some information. This count is also used as the base for determining the reliability

of the compensatory instruction, bilingual instruction, group size, and adult involved dimensions of the form.

One measure of reliability could be derived from the ratio of the number of matches to the total number of blocks used. It is the most stringent measure used in this study since an exact match between the teacher and observer responses is required before each observation is counted as correct.

The second measure of reliability is an adjusted measure which allows for slight differences in the times that the teacher and observer make their observations. The type of situation taken into account by this measure is illustrated in the following table.

	Time Block					
	1	2	3	4	5	6
Teacher	1	1	1	3	3	3
Observer	1	1	3	3	3	3

In time block 3 the activity changes. The teacher elected to code this activity "1" (i.e., Language Arts) since the activity continued into this time block. The observer, however, noted the change somewhere in the middle of the ten minute block and elected to code the new activity "3" (i.e., Mathematics). As in this example, such mismatches occur when activities change within blocks and are somewhat arbitrarily determined by the teachers' and observers' perceptions of how much of the ten minute block was actually devoted to each activity. The same pattern can also arise from slight differences in the clocks used by the two observers. Such errors could be discounted since the

error reflects a minor variation in time judgment rather than a misperception of the activity.

In order to account for these "shift" errors, the mismatch was discounted if either observation seemed to be a continuation of a previous activity. Specifically, a cross-over indicator was computed for each non-matching block. This is given as:

$$CERR = (T_n - O_{n-1}) (O_n - T_{n-1})$$

where: T_n (O_n) indicates a teacher (observer) observation at time n.

In the example given above this becomes:

$$\begin{aligned} CERR &= (T_3 - O_2) (O_3 - T_2) \\ &= (1 - 1) (3 - 1) \\ &= 0 \end{aligned}$$

The indicator is 0 when a shift error has occurred. The number of times $CERR = 0$ is counted and added to the number of "true" matches to give a corrected count of matches. The ratio of this corrected count of matches to the total number of valid blocks gives the adjusted reliability estimate. This adjustment occurs in a number of situations that differ from that of the table above. Some of the major types are given in Table 1.

Pattern I is included to show that perfect match results in no corrections, since CERR is not computed unless there is a mismatch. In pattern II the observation at Time 5 is in question. Using the

shift correction gives the teacher and/or observer the benefit of the doubt and counts this mismatch as a timing problem rather than an error of perception. Pattern III demonstrates that brief activities are effectively picked up if they are matched. However, if they are mismatched, the algorithm is fooled and the error is incorrectly discounted. This is shown in patterns IV and V.

Table 1

		Time								Matches	Shift Corrections
		1	2	3	4	5	6	7	8		
I	Teacher	1	1	1	1	1	2	2	2	8	0
	Observer	1	1	1	1	1	2	2	2		
II	Teacher	1	1	1	1	2	2	2	2	4	1
	Observer	1	1	1	1	1	3	3	3		
III	Teacher	1	1	1	1	2	1	1	1	6	2
	Observer	1	1	1	1	1	2	1	1		
IV	Teacher	1	1	2	1	1	1	1	1	6	2
	Observer	1	1	1	3	1	1	1	1		
V	Teacher	1	1	2	1	1	1	1	1	0	2
	Observer	2	2	1	2	2	2	2	2		

A more complicated procedure could, no doubt, be devised to handle the special cases of patterns IV and V. However, the current procedure sufficed for our purposes. The direct match measure was used to compute a lower bound on the reliability and the corrected measure was used to compute an upper bound. The two measures are given as:

$$R_L = \# \text{ matches} / \# \text{ valid blocks}$$

$$R_U = (\# \text{ matches} + \# \text{ shift corrections}) / \# \text{ valid blocks}$$

The analysis consisted basically of an examination of the distribution of these reliability estimates over the sample of 124 validation records. Based on previous experience, the following judgments were used to evaluate the reliabilities:

Range

90 - 100	Excellent
80 - 90	Very Good
70 - 80	Good
60 - 70	Acceptable
0 - 60	Poor

More detailed analyses were attempted in order to identify some of the possible sources of error which might be affecting the reliability estimates. Internal sources of error were examined by counting the number of mismatches of various types. This is basically a "confusion" analysis which served to identify code categories which were indistinguishable by the teacher and the observer. Some likely places for this confusion to occur are in the management aspects of classroom activities, e.g., lunch, recess, and classroom management, and in adjacent group sizes, especially between "10-20" and "20 and above" categories. Adult involved was also checked since the observer did not know all the personnel in the classroom and was, therefore, likely to make more errors in this section.

Results

The means of the direct and corrected matching measures for the five dimensions of the CAL are presented for each site in Table 2. The means are generally in the good to very good range for content area and in the excellent range for the compensatory services and bilingual sections. For the group size and adult involved dimensions, the means drop into the acceptable range for the direct measures, while the means for the corrected measures are still in the good to very good range.

The surprisingly high means for the compensatory services and bilingual instruction areas are due to the rarity of these events in the classroom. Thus, the "matches" used in the reliability measure are based mostly on non-responses to these areas and are artificially high. For compensatory services, 75% of the cases reported no services and 98% of the cases reported no bilingual instruction.

Since only 3 of 124 cases reported any bilingual instruction, no further analysis of the data can be supported. For the 34 cases that reported some compensatory services, matching measures were computed using a new base. This new base was a count of the number of blocks where either the teacher or the observer reported that compensatory services were being delivered. The size of the base ranged from 1 to 19 with a median of 3. With such small numbers to use as baselines, the resulting reliability estimates will be very sensitive to the number of matches found between teacher and observer. For instance, with a base of 4, two matches give a reliability estimate of .50 while three give an estimate of .75. A change of one match

Table 2

Means of Direct and Corrected Match Measures by Site

Site	N	Content Area		Compensatory Direct*	Bilingual Direct*	Group Size		Adult Involved	
		Direct	Corrected			Direct	Corrected	Direct	Corrected
1	8	.701	.817	1.000	1.000	.691	.797	.610	.753
2	7	.763	.849	.976	.981	.721	.823	.628	.757
3	8	.714	.818	.987	1.000	.472	.579	.557	.710
4	6	.532	.775	.995	1.000	.550	.585	.523	.567
5	9	.642	.757	.985	1.000	.695	.818	.573	.746
6	8	.770	.902	.990	1.000	.653	.760	.633	.758
7	9	.655	.764	.997	1.000	.799	.880	.636	.757
8	17	.773	.843	.968	1.000	.765	.826	.751	.836
9	10	.757	.849	.991	.991	.759	.840	.666	.780
10	6	.669	.805	.987	1.000	.641	.776	.524	.660
11	7	.758	.828	.996	1.000	.829	.902	.815	.888
12	9	.728	.810	.994	1.000	.578	.683	.518	.651
13	9	.733	.826	.985	1.000	.701	.786	.646	.749
14	7	.720	.852	.974	.985	.664	.797	.535	.662
15	4	.716	.815	1.000	1.000	.634	.727	.720	.817
Total	124	.716	.822	.987	.997	.688	.782	.631	.749

*Corrected means are not presented for the Compensatory and Bilingual Areas since little significant change was made by the correction.

moves the estimate from the "poor" level to the "good" level. Thus, this sub-analysis should be interpreted with caution. Seemingly low reliabilities may not, in fact, be very bad.

The means for the revised estimates are given in Table 3. This table includes means of the base for each site to help in interpreting the reliability estimates. Taking the previous caution into consideration, the results for the corrected measures in particular are generally good, except for site 13. In the latter case, the estimate is based on one case and should not be taken as a general indication of poor reliability.

Since the means for the content area section were sufficiently high, they were not considered further. The size and adult sections, however, had generally lower means and were subjected to more detailed analyses.

Taking the size variable first, a "confusion" matrix is presented in Table 4. The entries in this table are the percentages of the total validation sample ($N = 124$) reporting more than one mismatch between the teacher (column) category and the observer (row) category. For example, the figure "7.3" in row 1, column 2, means that in 9 of the 124 cases an observer coded 1 ("just this student") and a teacher coded 2 ("small group 2-5") for more than 1 block. A one block leeway is given since a single discrepancy might be due to a shift error. The larger percentages are in the column and row corresponding to large groups (code 5). The confusion with individual instruction (code 1) is especially troublesome. If it is assumed that the teacher is basing the judgment on the actual task the pupil is performing, then the

pattern of Table 3 suggests that the observer, not having access to this information, is responding only to the size of the group containing the child. This would explain the high percentage of cases with mismatches (13.7%) indicated in row 5, column 1, of Table 3. The higher percentage (33.9%) in row 1, column 5, of the table is harder to explain. One hypothesis is that teachers felt they were dealing with large groups even though they gave occasional aid to individual students. This would occur, for instance, in free study periods, where a teacher might provide occasional individual attention to a few students.

Table 3
Means of Matching Measures for Compensatory Services
Using Adjusted Bases

Site	N	Direct	Corrected	Base
1	0	----	----	----
2	3	.912	.982	8.33
3	3	.583	.778	3.67
4	1	.800	1.000	5.00
5	2	.400	.625	4.50
6	3	.619	.833	4.33
7	1	.667	1.000	3.00
8	6	.539	.725	7.00
9	4	.500	.875	2.00
10	2	.500	.667	3.00
11	2	.833	1.000	5.00
12	2	.500	.750	3.00
13	1	.000	.167	6.00
14	4	.450	.800	4.00
15	-	----	----	----

A second observation is that there seems to be confusion between adjacent categories. Note especially the percentages which are blocked in Table 4. These percentages, e.g., 21% for row 4, column 5, are generally higher than the rest. The inference is that the actual group size was close to the cutoff which distinguished one code from another, allowing for potential errors in estimating group sizes. This inference is strengthened by the occurrence of higher percentages for larger group sizes, i.e., 18.5 and 21, where the actual group size is more difficult to judge. While it would seem that a simple head count would have obviated the problem, the resulting error is not severe and could be discounted.

Table 4
Percentages of Sample Reporting More than One
Confusion of Group Size

Observer	Teacher					
	1	2	3	4	5	No response
1	x	7.3	8.1	15.3	33.9	2.4
2	8.1	x	5.6	4.0	12.1	3.2
3	5.6	1.6	x	9.7	9.7	2.4
4	6.5	1.6	6.5	x	21.0	4.0
5	13.7	6.5	7.3	18.5	x	25.0
No Response	5.6	0.8	0.8	4.0	16.9	x

Allowing for this type of inaccuracy, a new matching measure could be computed, which allowed the teacher and observer to differ as long as the judgments were not too difficult. A judgment was defined as "not too different" if adjacent group sizes were coded by

the teacher and observer. In these cases, a match was counted. The results are shown in Table 5. As can be seen, the direct and corrected accuracy measures both increased sharply.

Table 5

Adjusted Matching Measures for Group Size

Site	Direct	Corrected
1	.753	.853
2	.804	.869
3	.652	.735
4	.669	.698
5	.812	.905
6	.794	.878
7	.841	.908
8	.829	.882
9	.810	.886
10	.742	.842
11	.876	.930
12	.691	.756
13	.750	.827
14	.782	.885
15	.700	.773
Average	.775	.850

The implications of this analysis are two-fold. First, the teacher seems to err on the side of larger group sizes, suggesting that (s)he is reporting the planned structure of the activity rather than a particular interaction with a pupil. Thus, a large group activity would be coded as such even though attention might be given to an individual during this activity. Second, there is some ambiguity about the actual sizes of the groups involved, especially those relating to adjacent categories. This, for the most part, is something that the analysis must live with. One can, perhaps, meet the problem

directly by collapsing categories, especially 4 and 5 where the problem is greatest.

A similar confusion analysis is displayed in Table 6, for the type of adult involved. Here the major finding is that observers tend to use the no adult and no response categories more than do teachers. Note, however, that the confusion between homeroom teacher and no adult seems to occur in both directions, as can be seen in row 9, column 1, and row 1, column 9, of Table 6. Again this may be due to teachers' basing their judgments on what the child is actually working on rather than on what his/her apparent activities are.

The observers, who don't know what the child is actually doing, elect to code a no response or no adult. This would occur, for instance, in a supervised independent activity, where the teacher feels that, even though there is no interaction with the pupil, there is an adult involved. It may be that the teacher or other adult made the assignment, or occasionally provides assistance, so the teacher feels (s)he or that other person is the responsible adult. That the reverse situation occurs, i.e., that the observer reports that the teacher is involved while the teacher reports that no adult is involved, could be explained in the same way by simply reversing the teacher and observer perceptions in the above discussion. This is not very satisfying since looking at the no adult row of Table 6, the observer also codes no adult when the teacher codes some of the other adult involved categories, while the reverse is not always true.

For instance, the observer codes no adult when the teacher codes instructional specialist in 6.5% of the cases, while there are

Table 6

Percentages of Sample Reporting More Than One Confusion on Adult Involved

Observer	Teacher									
	1	2	3	4	5	6	7	8	9	No Response
1. Homeroom Teacher	x	8.1	1.6	0.0	0.0	0.8	1.6	0.0	32.3	11.3
2. Another Classroom Teacher	4.8	x	4.8	0.0	2.4	0.0	0.8	0.0	9.7	4.0
3. Instructional Specialist	4.8	9.1	x	1.6	0.0	0.0	0.8	0.0	0.8	2.4
4. Student Teacher	0.0	0.0	1.6	x	0.8	0.0	0.0	0.0	1.6	0.0
5. Aide, Assistant, or Adult Tutor	5.6	1.6	0.8	0.8	x	0.0	0.8	0.0	4.8	4.0
6. Parent/Other Adult	0.0	0.0	0.8	0.0	1.6	x	0.0	0.0	0.0	0.8
7. Other Adult	1.6	0.0	4.0	0.0	0.8	0.8	x	0.0	1.6	0.8
8. Older Student/Peer Tutor	0.0	0.0	0.0	0.0	0.0	0.0	0.0	x	0.8	0.0
9. No Adult Involved Directly	66.9	21.0	6.5	0.8	9.1	1.6	4.8	0.0	x	4.0
10. No Response	18.5	2.4	0.8	0.8	1.6	0.0	0.0	0.0	6.5	x

no cases where the teacher codes no adult and the observer codes instructional specialist. Since observers follow the child from activity to activity, it is possible that they are more aware of the child's experiences in non-homeroom situations. While this would explain the source of error, it does pose some problems for analysis. Unless one assumes that the adult responsible for the pupil is important, regardless of direct involvement with the child, this source of confusion could affect the interpretation of the result.

In addition to this confusion in the usage of the no adult category, there seems to be some confusion among homeroom teacher, other teacher, and instructional specialist as indicated by the blocked area of Table 6. This could mean that two adults were involved and that the teacher and observer had different opinions about who was in charge. It could also mean that situations occurred where the observer did not know who the real homeroom teacher was. In neither case is the problem serious, if it is presumed that the teacher, who knows the lesson plans, is in a position to make the better judgment.

Since there is no substantive justification for collapsing certain of the categories of adult involved, no adjustments were made to the matching estimates. In fact, it is necessary to treat the teacher and specialist categories separately for the analysis. Collapsing these, as suggested by the confusion matrix, would be in violation of the research objectives of the study.

Conclusion

Reliability estimates based on counts of matches between teacher and observer ratings indicate that the Classroom Activities Log is capable of providing good estimates of the quantity of instruction delivered in various content areas. When corrections are made for minor errors in judgment by either teachers or observers, this reliability increases to what we consider the very good range.

Uncorrected reliability estimates for the "Group Size" and "Adult Involved" dimensions of the CAL, measures of what we call the "intensity" of the instruction, are in the acceptable range; corrections bring both of these up to the good level. Lower reliability for the "Group Size" and "Adult Involved" dimensions could be associated with differences in the amount of information available to teachers and observers, and in the case of group size, to the ambiguity arising from similarity of adjacent categories. These results are very encouraging, especially if we choose to believe that teacher reports are more accurate or if we are able to collapse certain categories.

In addition to reliability estimates, analyses of data collected on more than 7,000 students from more than 3,800 teachers reveal variation in quantity and intensity of instruction within school districts over time and between various types of students. These findings suggest a certain degree of construct validity for the scale. Preliminary comparisons of CAL data with other programmatic data available within these districts provide some concurrent validity as well. These findings, together with other evidence supporting the conclusion that the CAL can collect quantity of instruction data in an efficient and cost-effective manner, point to the present and future value of the instrument.

References

- Berliner, D. C. Changing Academic Learning Time: Clinical Interventions in Four Classrooms. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, April 1978.
- Bloom, B. S. Human Characteristics and School Learning. New York: McGraw-Hill, 1976.
- Carroll, J. B. A model for school learning. Teachers College Record, 1963, 64, 723-733.
- Classroom Activities Log (CAL) Instructions Manual. Cambridge, MA: Abt Associates, Inc., February 1977.
- Coleman, J. S., et al. Equality of Educational Opportunity. Washington, D.C.: U. S. Government Printing Office, 1966.
- Fisher, C. W. Teaching Behaviors, Academic Learning Time and Student Achievement: An Overview of Phase III-B of the Beginning Teacher Evaluation Study. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, April 1978.
- Fisher, C. W., Filby, N. N. and Marliave, R. S. Instructional Time and Student Achievement in Second Grade Reading and Mathematics. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, April 1978.
- Frechtling, J. A. and Nyitray, M. S. The Structure and Content of Compensatory Education Programs: A Research Strategy for Evaluating Title I's Effects on Services and Effects on Students. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, April 1978.
- Harnischfeger, A., and Wiley, D. E. Conceptual Issues in Models of School Learning, Studies of Educative Processes, No. 10. Chicago: CEMREL, Inc., May 1977.
- Karweit, N. Exposure to Schooling: Method, Conclusion, Policy. Educational Researcher, February 1976, 5, 15-17.
- Marliave, R. Academic Learning Time and Achievement: The Validation of a Measure of Ongoing Student Engagement and Task Difficulty. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, April 1978.
- Jencks, C. S., et al. Inequality: A Reassessment of the Effect of Family and Schooling in America. New York: Basic Books, 1972.

Mosteller, F., and Moynihan, D. P. (Eds.) On Equality of Educational Opportunity.
New York: Random House, 1972.

Rosenshine, B. and Furst, N. The use of direct observation to study teaching.
In Robert Travers (Ed.), Second Handbook of Research on Teaching.
Chicago: Rand McNally, 1973.

Vanecko, J. J., Archambault, F. X., and Ames, N. L. ESEA Title I Allocation
Policy: Demonstration Study—Analysis of Baseline Data. Cambridge,
MA: Abt Associates Inc., 1977.

Wiley, D. E., and Harnischfeger, A. Explosion of a myth: Quantity of schooling
and exposure to instruction, major educational vehicles. Educational
Researcher, April 1974, 3, 7-12.